

А.С. БАБІЙ, О.Ф. ЛАНОВИЙ, канд. техн. наук

СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ЗЛОЧИННОСТІ

В статті пропонується метод побудови моделей, в яких враховуються суттєві фактори і не враховані неістотні, що визначають злочинність в конкретний інтервал часу в окремому регіоні. Використовуючи доступні статистичні дані по регіонах в статті приведені приклади використання поданих моделей.

Якщо взяти за основу постулат, згідно якого злочинність є соціальним явищем і не може розвиватися окремо від суспільства, тоді будь-які зміни в економічній, політичній, соціальній сфері спричиняють певну реакцію в суспільстві, в тому числі і кримінальній складовій.

Така ідея закладена в кримінальному та кримінально-процесуальному кодексах, де сказано про обов'язок слідчого виявити умови, які сприяють скоєнню злочину, а також необхідність врахування висновків правоохоронних установ органами державної влади [1].

Для реалізації цих функцій на регіональних рівнях в системі правоохоронних органів створено підрозділи, які працюють над визначенням розвитку злочинності і створенням прогнозів.

Стан злочинності залежить від багатьох факторів які визначають суспільне життя. В різні моменти часу та на різних територіях одні й ті ж фактори впливають по різному на кримінальну ситуацію. Тому однією з найважливіших проблем аналітичного дослідження кримінологічної складової є визначення суттєво впливаючих факторів для побудови моделі злочинності. Від цього залежить складність і точність створеної моделі і відповідно якість висновків експерта про розвиток злочинності в конкретному регіоні.

Створенню методу для побудови моделей, в яких враховуються суттєві фактори і не враховані неістотні, що визначають злочинність в конкретний інтервал часу в окремому регіоні, присвячена дана робота.

В органах внутрішніх справ ведуться обліки злочинів і злочинців, які можуть бути використаними для аналізу злочинності, зокрема для побудови вказаних моделей [2]. Облікові дані можна подати у вигляді динамічного ряду:

$$Y = (Y_1, Y_2, \dots, Y_k),$$

де $Y_i, i = \overline{1..n}$ - сумарна кількість скоєних злочинів, або окремих видів для i -го моменту часу.

Позначимо сукупність факторів, які впливають на злочинність:

$$\vec{x} = (x_1, x_2, \dots, x_k),$$

де k – кількість факторів. Кожний з факторів приймає декілька значень, тому сукупність факторів (факторний простір) можна подати у матричній формі:

$$X = \{x_{ij}\}, \text{ де } i = \overline{1, n}, j = \overline{1, k}.$$

Серед множини факторів існують такі, що істотно впливають на зміну значення функції, і такі, що впливають несуттєво. Необхідно знайти такий вектор $\vec{x}' = (x_1, x_2, \dots, x_m), m < k$, варіабельність якого б вичерпував всю, або мала б максимально можливу варіабельність початкових показників стосовно результуючої функції. Цього можна досягнути, використовуючи кореляційний аналіз.

Кореляційна залежність визначається різними параметрами, серед яких найбільше поширення одержали показники, які характеризують взаємозв'язок двох випадкових величин (парні показники): кореляційний момент, коефіцієнт кореляції.

Оцінка кореляційного моменту[3] (коефіцієнта коваріації) двох варіант x_j і x_k обчислюється за початковою матрицею X :

$$\xi_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \mu_1(x_j))(x_{ik} - \mu_1(x_k))$$

Цей показник незручний для практичного застосування, оскільки має розмір, який дорівнює добутку розмірів варіант і за його величиною важко судити про залежність параметрів.

Тому для оцінки залежності двох випадкових величин використовується коефіцієнт кореляції r_{jk} , який для нормованих випадкових величин дорівнює:

$$\rho_{jk} = \frac{1}{n} \sum_{i=1}^n u_{ij} u_{jk}$$

Перехід до нормованих випадкових величин здійснюється за формулами:

$$u_{ij} = (x_{ij} - \mu_{1j}) / \sigma_j,$$

де μ_{1j} - вибіркова середня, σ_j - середнє квадратичне відхилення для кожного фактора.

Коефіцієнт ρ_{jk} характеризує значимість лінійного зв'язку між випадковими величинами (параметрами):

- при $\rho_{jk} = 1$ значення u_{ij} і u_{ik} збігаються, тобто значення параметрів приймають однакові значення. Інакше кажучи, має місце функціональна залежність: знаючи значення одного параметра, можна однозначно вказати значення іншого параметра[4];

- при $\rho_{jk} = -1$ величини u_{ij} і u_{ik} приймають протилежні значення. І в цьому випадку має місце функціональна залежність;

- при $\rho_{jk} = 0$ величини u_{ij} і u_{ik} практично не зв'язані одна із одною лінійним співвідношенням. Це не означає відсутності інших (наприклад, нелінійних) зв'язків між параметрами;

- при $|\rho_{jk}| > 0$ і $|\rho_{jk}| < 1$ однозначного лінійного зв'язку величин u_{ij} і u_{ik} немає. І чим менша абсолютна величина коефіцієнта кореляції, тим меншою мірою за значеннями одного параметра можна знайти значення іншого.

Використовуючи поняття коефіцієнта кореляції, матриці експериментальних даних X можна поставити у відповідність квадратну матрицю оцінок коефіцієнтів кореляції (кореляційну матрицю):

$$\rho = \begin{vmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1m} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2m} \\ \dots & \dots & \dots & \dots \\ \rho_{m1} & \rho_{m2} & \dots & \rho_{mm} \end{vmatrix}$$

Оцінка коефіцієнта кореляції, знайдена за обмеженою вибіркою, практично завжди відрізняється від нуля. Але з цього не випливає, що коефіцієнт кореляції генеральної сукупності також відмінний від нуля. Потрібно оцінити значимість вибіркової величини коефіцієнта кореляції або, відповідно до постановки задач перевірки статистичних гіпотез, перевірити гіпотезу про рівність нулю коефіцієнта кореляції. Якщо гіпотеза H_0 про рівність нулю коефіцієнта кореляції відхиляється, то вибіркового коефіцієнта значимий, а відповідні величини зв'язані лінійним співвідношенням. Якщо гіпотеза H_0 приймається, то оцінка коефіцієнта не значима, і величини лінійно не зв'язані одна із одною (якщо з інших міркувань фактори можуть бути зв'язаними, то краще говорити про те, що за наявними експериментальними даними цей взаємозв'язок не встановлений).

За критерій перевірки нульової гіпотези H_0 застосовується випадкова величина

$$t = \frac{|\rho_{ik}| \sqrt{n-2}}{\sqrt{1-\rho_{ik}^2}}$$

Якщо модуль коефіцієнта кореляції відносно далекий від одиниці, то величина t при справедливості нульової гіпотези розподілена за законом Стюдента з $n-2$ степенями свободи. Конкуруюча гіпотеза H_1 відповідає твердженню, що значення ρ_{jk} не дорівнює нулю (більше або менше нуля).

Перевірка гіпотези H_0 про рівність нулю коефіцієнта парної кореляції двовимірної нормально розподіленої випадкової величини здійснюється в такій послідовності:

- обчислюється значення статистики t ;
- при рівні значимості α для двосторонньої критичної області визначається критична точка розподілу Стюдента $t_{кр}(n-2; \alpha)$;
- порівнюється значення статистики t із критичним значенням $t_{кр}(n-2; \alpha)$. Якщо $t < t_{кр}(n-2; \alpha)$, то немає підстав відхиляти нульову гіпотезу, інакше гіпотеза H_0 відхиляється (коефіцієнт кореляції значимий).

Коли модуль величини ρ_{ik} близький до одиниці, розподіл ρ_{ik} відрізняється від розподілу Стюдента, оскільки значення $|\rho_{ik}|$ обмежено праворуч одиницею. В цьому випадку застосовується перетворення

$$y_{ik} = 0,5 \ln \frac{1 + |\rho_{ik}|}{1 - |\rho_{ik}|}$$

Величина y_{ik} не має зазначеного обмеження, вона при $n > 10$ розподілена приблизно нормально з центром

$$\mu_1(\rho_{ik}) = 0,5 \ln \frac{1 + |\rho_{ik}|}{1 - |\rho_{ik}|} + 0,5 \frac{|\rho_{ik}|}{n-1}$$

і дисперсією

$$\mu_2(\rho_{ik}) = s_2(\rho_{ik}) = \frac{1}{n-3}$$

Розглянемо приклади застосування кореляційного аналізу для розв'язання задач ОВС. Дані (кількість злочинів) в умовних одиницях на 10 тис. населення для Донецької області подані поквартально на графіку (рис.1). Інші дані подаються в табл. 1 [5, 6].

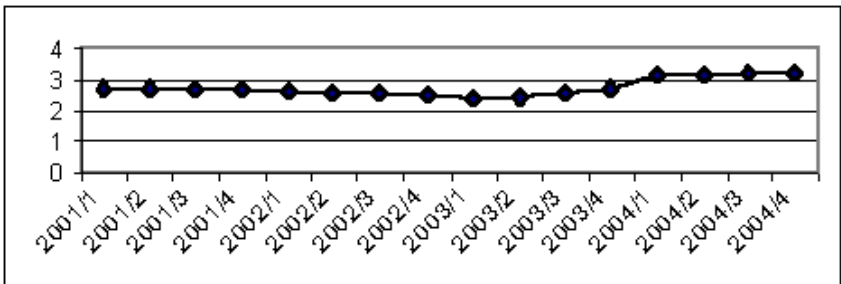


Рис. 1. Динаміка злочинності в Донецькій області.

Таблиця 1.
Статистичні дані (Донецька область)

	2000	2001	2002	2003
Коеф. мігр. приросту	-1,3	-1,7	-2	-1
Шлюбність	5	5,9	6,1	7,2
Розлучуваність	4,7	4,1	4,1	3,7
Кількість зайнятого населення	2176,5	2101,4	2103,9	2129
Безробіття	231,7	241,6	218,3	183,9
Не зайняті трудовою діяльністю	97,9	74,9	78,4	72,9
Вимушена неповна зайнятість	13,5	6,6	3,8	1,9
Неповний робочий день	10,1	8,1	9	6,4
Сер. зар. плата	292	383	452	550
Житловий фонд	98614	99447	100504	100630
Кількість студентів	350	386	410	441
Кількість ВНЗ	94	93	93	93

Утворимо початкову матрицю $X=(x_{ij})$, яка складається з місячних відомостей про кількість злочинів, та інших даних, приведених у табл. 1. У цій матриці рядок i визначає фактор, а стовпчик j визначає значення фактора в j -му місяці. Матриця X нормується. Оцінки коефіцієнтів кореляції (табл. 2) знаходяться зі співвідношень:

$$\rho_{ik} = \frac{1}{12} \sum_{i=1}^{12} u_{i1} u_{ki} .$$

Таблиця 2.
Значення коефіцієнтів кореляції і критерію Стьюдента.

	Коеф. корр	t-критерій
Коеф. мігр. приросту	0,955695	2,29577061
Шлюбність	-0,84266	1,10662442
Розлучуваність	0,738408	0,77427099
Кількість зайнятого населення	0,718329	0,73010316
Безробіття	0,735633	0,76791502
Не зайняті трудовою діяльністю	0,635783	0,58243957
Вимушена неповна зайнятість	0,897888	1,44222769
Неповний робочий день	0,356742	0,27002152
Сер. зар. плата	-0,9566	2,32110007
Житловий фонд	-0,98904	4,73725281
Кількість студентів	-0,94557	2,05465668
Кількість ВНЗ	0,738408	0,77427099

Критичне значення критерію Стюдента при $\alpha = 0,15$ дорівнює 1,46. Статистика критерію більша за критичне значення для таких факторів:

- Житлового фонду
- Середньої заробітної плати
- Міграційного приросту населення.
- Кількість студентів.

Таким чином, при побудові моделі злочинності в даному регіоні вищевказані фактори є значущими (коefficient кореляції не дорівнює нулю), а інші коefficientи при факторах варто визнати рівними нулю.

Кримінологічні дослідження також відносять вищевказані фактори до числа тих, що найбільше впливають на злочинність [1].

Розглянемо аналогічні статистичні дані по Вінницькій області (рис. 2, табл. 3) [5, 6].

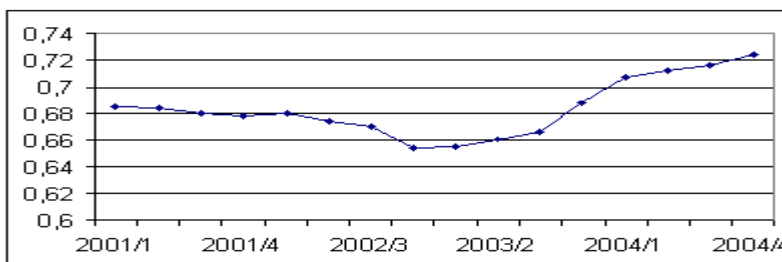


Рис. 2. Динаміка злочинності в Донецькій області.

Таблиця 3.

Статистичні дані (Вінницька область)

	2000	2001	2002	2003
Чисельність населення	1811	1791	1772,4	1753,9
Коеф. мігр. приросту	-2,6	-3,4	-1,1	-1,6
Шлюбність	5,8	6,3	6,6	7,6
Розлучуваність	3,3	3,3	3,6	3,5
Кількість зайнятого населення	833	792,1	799	826,7
Безробіття	83,4	85,7	71	49,2
Не зайняті трудовою діяльністю	40,3	41,6	36,7	31,8
Вимушена неповна зайнятість	19,2	8,1	6,1	3,5
Неповний робочий день	10,9	14,2	15,9	16,1
Сер. зар. плата	159	215	265	334
Житловий фонд	42994	43116	43486	43363
Кількість студентів	270	291	310	399
Кількість ВНЗ	31	31	31	33

Оцінки коефіцієнтів кореляції приведені в табл. 4.

Таблиця 4.

Значення коефіцієнтів кореляції і критерію Стьюдента.

	Коеф. корр	t-критерій
Коеф. мігр. приросту	-0,86125	1,19838567
Шлюбність	-0,88536	1,34658305
Розлучуваність	-0,98323	3,81197189
Кількість зайнятого населення	0,521649	0,43234751
Безробіття	0,946247	2,06866662
Не зайняті трудовою діяльністю	0,903741	1,49281395
Вимушена неповна зайнятість	0,750745	0,80360783
Неповний робочий день	-0,86638	1,22678269
Сер. зар. плата	-0,93129	1,80775379
Житловий фонд	-0,99838	12,3947361
Кількість студентів	-0,93267	1,82825211
Кількість ВНЗ	0,0001	0,10212201

Критичне значення критерію Стьюдента при $\alpha = 0,15$ дорівнює 1,46.

В цьому випадку статистика критерію більша за критичне значення для таких факторів:

- Житлового фонду
- Розлучуваність
- Безробіття
- Кількість не зайнятих трудовою діяльністю
- Середня заробітня плата
- Кількість студентів

Відповідно, можна зробити висновки про те, що в данному випадку при побудові моделі необхідно врахувати ці фактори як значущі, а інші визнати такими, що не здійснюють суттєвого впливу на злочинність в данному регіоні.

Список літератури: 1. Иншаков С.М. Криминология, Учеб.- М.: Юриспруденция, 2000. 426с. 2. Бодянский Е.В., Иващенко П.А., Яковлев С.В., Червяков И.М. Правовая статистика. Харьков, УВД. 1996. 98 с. 3. Дронов С.В. Многомерный статистический анализ, Барнаул: изд-во алтайского гос. университета, 2003, 213 с. 4. Булашев С.В. „Статистика” М.: Компания Спутник+ , 2003, 245с. 5. Статистичний бюлетень мін. статистики України // 2003-2005р. 6. Статистичний довідник „Регіони України” К.: 2000, т.1, 546 с.

Надійшла до редколегії 10.05.06